

Aberystwyth University

Improving fuzzy rule interpolation performance with information gain-guided antecedent weighting

Li, Fangyi; Li, Ying; Shang, Changjing; Shen, Qiang

Published in:
Soft Computing

DOI:
[10.1007/s00500-017-2805-2](https://doi.org/10.1007/s00500-017-2805-2)

Publication date:
2018

Citation for published version (APA):

Li, F., Li, Y., Shang, C., & Shen, Q. (2018). Improving fuzzy rule interpolation performance with information gain-guided antecedent weighting. *Soft Computing*, 22(10), 3125-3139. <https://doi.org/10.1007/s00500-017-2805-2>

Document License CC BY

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Improving fuzzy rule interpolation performance with information gain-guided antecedent weighting

Fangyi Li^{1,2} · Ying Li¹ · Changjing Shang²  · Qiang Shen²

Published online: 4 September 2017

© The Author(s) 2017. This article is an open access publication

Abstract Fuzzy rule interpolation (FRI) makes inference possible when dealing with a sparse and imprecise rule base. However, the rule antecedents are commonly assumed to be of equal significance in most FRI approaches in the implementation of interpolation. This may lead to a poor performance of interpolative reasoning due to inaccurate or incorrect interpolated results. In order to improve the accuracy by minimising the disadvantage of the equal significance assumption, this paper presents a novel inference system where an information gain (IG)-guided fuzzy rule interpolation method is embedded. In particular, the rule antecedents in FRI are weighted using IG to evaluate the relative importance given the consequent for decision making. The computation of antecedent weights is enabled by introducing an innovative reverse engineering process that artificially converts fuzzy rules into training samples. The antecedent weighting scheme is integrated with scale and move transformation-based interpolation (though other FRI techniques may be improved in the same manner). An illus-

trative example is used to demonstrate the execution of the proposed approach, while systematic comparative experimental studies are reported to demonstrate the potential of the proposed work.

Keywords Fuzzy rule interpolation · Antecedent weighting · Reverse engineering

1 Introduction

Fuzzy set theory (Zadeh 1965) has gained rapid developments in a variety of scientific areas, including mathematics, engineering, and computer science. It has been successfully applied for many real-world problems, such as systems control, fault diagnosis and computer vision, as an effective tool to address the issues of imprecision and vagueness in modelling and reasoning. In particular, fuzzy expert systems have been developed using the idea of linguistic reasoning (also known as approximate reasoning), which reflects the way of cogitation of human beings and leads to new, more human, intelligent systems.

In general, an approximate reasoning system can be formalised as a fuzzy *if-then* rule-based inference mechanism that derives a conclusion given an input observation. Various techniques have been established to implement generalised *modus ponens* that facilitates reasoning when provided with imprecise inputs, mostly by following the basic idea of Compositional Rule of Inference (CRI) (Zadeh 1973). However, CRI is unable to draw a conclusion when a rule base is not dense but sparse. Sparse rule bases considered here are not referring to the quantity of rules in a given rule base, but the domain coverage of the antecedents of rules in the universe of discourse. That is, an input observation may have no over-

Communicated by P. Angelov, F. Chao.

✉ Changjing Shang
cns@aber.ac.uk

Fangyi Li
fal2@aber.ac.uk

Ying Li
lybyp@nwpu.edu.cn

Qiang Shen
qqs@aber.ac.uk

¹ School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, China

² Department of Computer Science, Institute of Maths, Physics and Computer Science, Aberystwyth University, Aberystwyth, UK

lap with any of the rules available and hence, no rule may be executed to derive the required consequent by applying CRI.

Fuzzy rule interpolation (FRI) (Kóczy and Hirota 1993a, b) plays a significant role in such sparse fuzzy rule-based reasoning systems. It addresses the limitation of conventional fuzzy reasoning that only uses CRI to perform inference, where the antecedents of all the rules within a given rule base cannot cover the whole problem domain. An estimation is able to be made by computing an interpolated consequent for the observation which has no rules matched.

A number of FRI methods have been proposed and improved in the literature (Hsiao et al. 1998; Chang et al. 2008; Huang and Shen 2006; Yang and Shen 2011; Yang et al. 2017; Jin et al. 2014). However, common approaches assume that the rule antecedents involved are of equal significance while searching for rules to implement interpolation. This can lead to inaccurate or incorrect interpolative results. This is because for many application of (fuzzy) decision systems, the decision is typically reached by an aggregation of conditional attributes, with each attribute making a generally different contribution to the decision making process. Weighted FRI methods (Diao et al. 2014) have therefore been introduced to remedy this equal significance assumption. For example, a heuristic method based on Genetic Algorithm is applied to learn the weights of rule antecedents (Chen and Chang 2011), but this leads to a substantial increase in computation overheads. An alternative work is to subjectively predefine the weights on the antecedents of the rules by experts, but this may restrict the adaptivity of the rules and, therefore, the flexibility of the resulting fuzzy system (Li et al. 2005).

In order to assess the relative significance of attributes with regard to the decision variable, information gain has been commonly utilised in data-driven learning algorithms (Mitchell 1997). By observing the property of information gains, this paper presents an innovative approach for rule interpolation. Information gain is integrated within an FRI process to estimate the relative importance of rule antecedents in a given rule base. The required information gains are estimated using an artificially generated decision table through a reverse engineering process which converts a given sparse rule base into a training data set. The proposed work helps minimise the disadvantage of the equal significance assumption made in common FRI techniques, thereby improving the performance of FRI. In particular, the paper presents an information gain-guided FRI method based on the popular scale and move transformation-based FRI (T-FRI) (Huang and Shen 2006). However, alternative FRI techniques may be employed for the same purpose if preferred.

The remainder of this paper is structured as follows. Section 2 outlines the background work that is required for the present development, including T-FRI, the basic concepts of information gain, and a simple iterative rule

induction method (for providing the initial rule base). Section 3 describes the proposed information gain-guided fuzzy rule interpolation approach, with a case study illustrating its execution process. Section 4 details the results of comparative experimental evaluations, supported by statistical tests and analysis. Finally, Sect. 5 concludes the paper and points out several further studies.

2 Background work

This section presents an overview of FRI based on scale and move transformations, a description of an iterative rule generation technique, and an outline of the concept of information gain.

2.1 Transformation-based FRI

An FRI system can be defined as a tuple $\langle R, Y \rangle$, where $R = \{r^1, r^2, \dots, r^N\}$ is a non-empty set of finite fuzzy rules (the rule base), and Y is a non-empty finite set of variables (interchangeably termed attributes). $Y = A \cup \{z\}$ where $A = \{a_j | j = 1, 2, \dots, m\}$ is the set of antecedent variables, and z is the consequent variable appearing in the rules. Without losing generality, a given rule $r^i \in R$ and an observation o^* can be expressed in the following format:

r^i : if a_1 is A_1^i and a_2 is A_2^i and \dots and a_m is A_m^i , then z is z^i

o^* : a_1 is A_1^* and a_2 is A_2^* and \dots and a_m is A_m^*
where A_j^i represents the value (or fuzzy set) of the antecedent variable a_j in the rule r^i , and z^i denotes the value of the consequent variable z in r^i .

A key concept used in T-FRI is the representative value $\text{Rep}(A_j)$ of a fuzzy set A_j , it captures important information such as the overall location in the domain of a fuzzy set and its shape. In general, given an arbitrary polygonal fuzzy set $A = (a_1, a_2, \dots, a_n)$ where $a_i, i = 1, 2, \dots, n$ denotes the vertex of the polygonal, its representative value $\text{Rep}(A)$ is defined by (Huang and Shen 2008):

$$\text{Rep}(A) = \sum_{i=1}^n w_i a_i \quad (1)$$

where w_i is the weight assigned to the vertex a_i . For simplicity, the weight of each vertex is typically assumed to be equal, i.e., $w_i = 1/n$.

Much research has adopted triangular membership functions to perform interpolation, which are the most commonly used in fuzzy systems. A triangular membership function is denoted in the form of $A_j = (a_{j1}, a_{j2}, a_{j3})$, where a_{j1}, a_{j3} represent the left and right extremities of the support (with membership values 0), and a_{j2} denotes the normal point (with

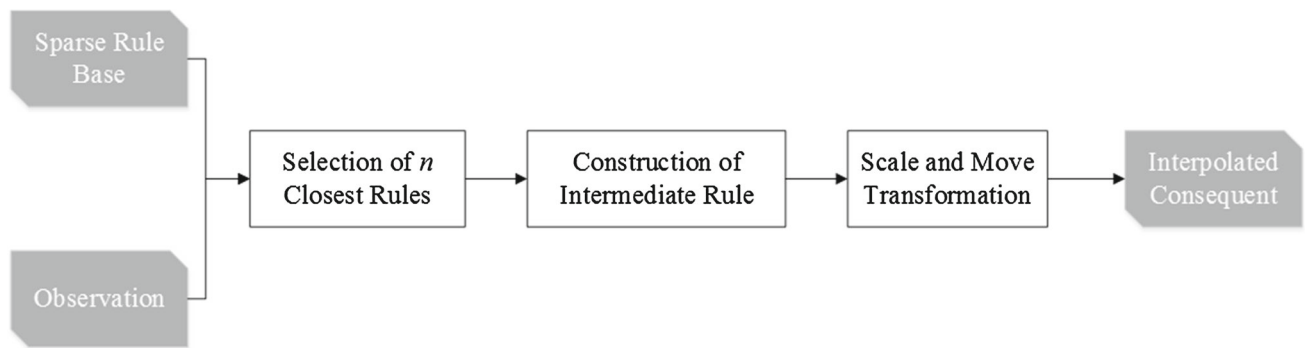


Fig. 1 Framework of transformation-based FRI

a membership value of 1). For such a fuzzy set A_j , $\text{Rep}(A_j)$ is defined as the centre of gravity of these three points:

$$\text{Rep}(A_j) = \frac{a_{j1} + a_{j2} + a_{j3}}{3} \quad (2)$$

The definition of representative values for more complex membership functions can be found in (Huang and Shen 2008).

Given a sparse rule base R and an observation o^* , as illustrated in Fig. 1, the T-FRI works as shown in Algorithm 1. This can be briefly described as follows.

Without being able to find a rule that directly matches the given observation, the closest rules to the observation are identified and selected instead. The selection criterion is based on the Euclidean distance metric (though other distance metrics may be considered for an alternative), which measures the similarity between the observation o^* and each rule r^p , $p = 1, 2, \dots, N$ in the sparse rule base. In general, the distance between an observation o^* and a rule r^q , or indeed between any two rules $r^p, r^q \in R$, is determined by computing the aggregated distances between all the corresponding values of the antecedent variables:

$$d(v, r^q) = \sqrt{\sum_{j=1}^m d(A_j^v, A_j^q)^2} \quad (3)$$

where v is o^* or r^p (so A_j^v is A_j^* or A_j^p), depending on whether the distance is between an observation and a rule or between two rules, and

$$d(A_j^v, A_j^q) = \frac{|\text{Rep}(A_j^v) - \text{Rep}(A_j^q)|}{\max_{A_j} - \min_{A_j}} \quad (4)$$

is the normalised result of the otherwise absolute distance measure, so that distances are compatible with each other over different variable domains. The \max_{A_j} and \min_{A_j} in the denominator specify the maximal and minimal value of the antecedent A_j in its domain, respectively. In general, they will not be identical so that the calculation of the normalised

Algorithm 1 Transformation-based FRI (T-FRI)

Input: Sparse rule base R , Observation o^* , Number of closest rules n
Output: Interpolated consequent z^*

-Selection of Closest Rules:

- 1: Calculate distance $d(o^*, r^p)$ between o^* and each rule r^p , $p = 1, 2, \dots, N$ in the sparse rule base R ;
- 2: Choose n rules which have minimal distances as the closest n rules;

-Construction of Intermediate Rule r' :

- 3: Assign normalised weights w_j^i , $i = 1, \dots, n$ to each the j th antecedent of the i th selected closest fuzzy rule;
- 4: Aggregate each corresponding weighted antecedents of the n rules to obtain the antecedent of intermediate rule A_j' , $j = 1, 2, \dots, m$;
- 5: Calculate weights to each consequent of the i th selected closest fuzzy rule w_z^i , which is the mean of the normalised weights associated with the antecedents w_j^i in each rule:

$$w_z^i = \frac{1}{m} \sum_{j=1}^m w_j^i$$

- 6: The consequent variable of the intermediate rule z' is constructed with the weighted aggregation of the consequent variables of n closest rules z^i , $i = 1, \dots, n$;

-Computation of Scale and Move Factors:

- 7: **for** Each antecedent **do**
- 8: Scale operation: from A_j' to \hat{A}_j' (denoting the scaled intermediate fuzzy set), in an effort to determine the scale rate s_{A_j} ;
- 9: Move operation: from \hat{A}_j' to A_j^* to obtain a move ratio m_{A_j} ;
- 10: **end for**

-Calculation of z^* via Scale and Move Transformation:

- 11: Compute transformation factors for z' by averaging the corresponding values:

$$s_z = \frac{1}{m} \sum_{j=1}^m s_{A_j} \quad m_z = \frac{1}{m} \sum_{j=1}^m m_{A_j}$$

- 12: Scale and Move the fuzzy set of the intermediate consequent z' with the above calculated factors, resulting in the required interpolated result z^* : $z^* = T(z', s_z, m_z)$;
- 13: **Return** z^*

distance between two antecedents [i.e., Eq. (4)] is valid mathematically. In the extreme case, however, the denominator may be zero, which indicates that all the antecedents in the domain of a_j are the same. In this case, the normalised dis-

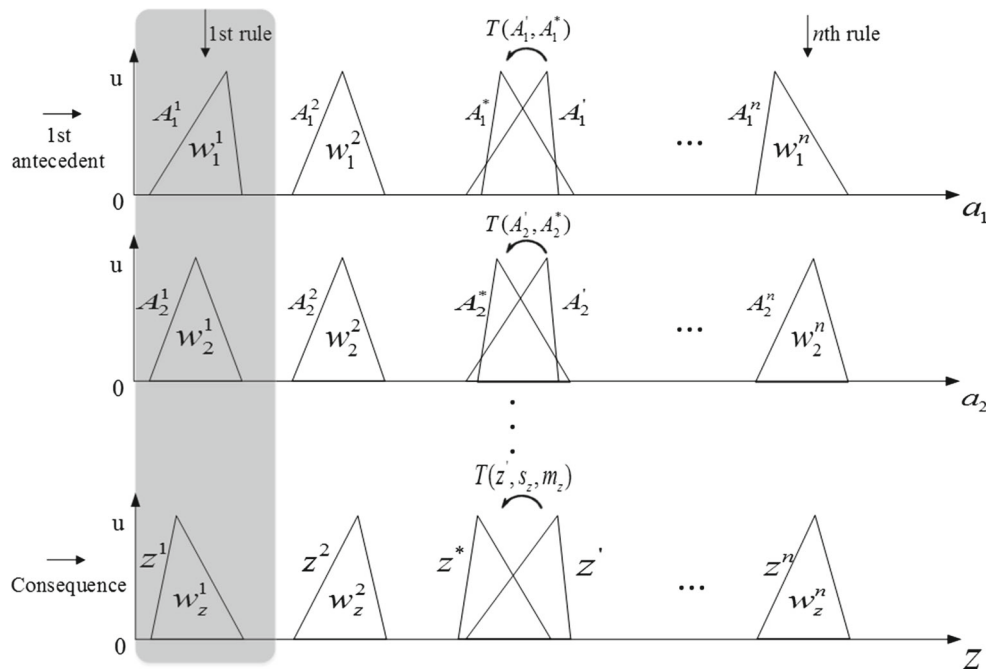


Fig. 2 Interpolation via scale and move transformations

tance is naturally defined to be zero (i.e., $d(A_j^v, A_j^q) = 0$ given that A_j^v always equals A_j^q).

Once the distances between a given observation and all rules in the rule base are calculated, the n rules which have minimal distances are chosen as the closest n rules with respect to the observation. In most applications of T-FRI, n is taken to be 2. The selection of the n closest rules sets up the basis upon which to construct a so-called intermediate rule r' . This construction process computes intermediate antecedent fuzzy sets A_j' , $j = 1, 2, \dots, m$, and an intermediate consequent fuzzy set z' , resulting in an artificially created rule:

$r' : \text{if } a_1 \text{ is } A_1' \text{ and } a_2 \text{ is } A_2' \text{ and } \dots \text{ and } a_m \text{ is } A_m', \text{ then } z \text{ is } z'$ which is in effect a weighted aggregation of the n selected closest rules.

Then, the antecedent values of the intermediate rule are transformed through a process of scale and move modification such that they become the corresponding parts of the observation, recording the transformation factors s_{A_j} and m_{A_j} , $j = 1, 2, \dots, m$ for each antecedent that are calculated. Finally, the interpolated consequent is obtained by applying the recorded factors to the consequent variable of the intermediate rule. This in effect implements fuzzy or generalised modus ponens.

The above process of scale and move transformations in an effort to interpolate the consequent variable can be summarised in Fig. 2, which can be collectively and concisely represented by: $z^* = T(z', s_z, m_z)$, highlighting the impor-

tance of the two key transformations required. The detailed computation involved in T-FRI can be referred to the original work (Huang and Shen 2006, 2008).

2.2 Information gain

Information gain has been widely adopted in the development of learning classifier algorithms, to measure how well a given attribute may separate the training examples according to the underlying classes (Mitchell 1997). It is defined via the entropy metric in information theory (Shannon 2001), which is commonly used to characterise the disorder or uncertainty of a system.

Formally, let $\mathbf{O} = (O, p)$ be a discrete probability space, where $O = \{o_1, o_2, \dots, o_n\}$ is a finite set of domain objects, with each having the probability p_i , $i = 1, \dots, n$. Then, the Shannon entropy of O is defined by

$$\text{Entropy}(O) = - \sum_{i=1}^n p_i \log_2 p_i \quad (5)$$

Regarding the task of classification, o_i , $i = 1, \dots, n$ represents a certain object, and p_i is the proportion of O which is labelled as the class j , $j = 1, \dots, m$, $m \leq n$. Note that the entropy is at its minimum (i.e., $\text{Entropy}(O) = 0$) if all elements of O belong to the same class (with $0 \log_2 0 = 0$ defined), and the entropy reaches its peak point (i.e., $\text{Entropy}(O) = \log_2 n$) if the probability of each category is equal; otherwise, the entropy is between 0 and $\log_2 n$.

Intuitively, the less the entropy value, the easier the classification problem. It is based on this observation that information gain has been introduced to measure the expected reduction in entropy caused by partitioning the values of an attribute. This leads to the popular decision tree learning methods (Quinlan 1986). Given a collection of examples $U = \{O, A\}$, $o_i \in O$ ($i = 1, \dots, n$) is an object which is represented with a group of attribute $A = \{a_1, \dots, a_l\}$ and a class label m . Information gain upon a particular attribute a_k , $k \in \{1, \dots, l\}$, is defined as

$$IG(O, a_k) = \text{Entropy}(O) - \sum_{v \in \text{Value}(a_k)} \frac{|O_v|}{|O|} \text{Entropy}(O_v) \quad (6)$$

where $\text{Value}(a_k)$ is the set of all possible values for the attribute a_k , O_v is the subset of O where the value of the attribute a_k is equal to v (i.e., $O_v = \{o \in O | a_k(o) = v\}$), and $|\cdot|$ denotes the cardinality of a set.

From the perspective of entropy evaluation over U , the second part of Eq. (6) shows that the entropy is measured via weighted entropies that are calculated over the partition of O using the attribute a_k . The bigger the value of information gain $IG(O, a_k)$, the better the partitioning of the given examples with a_k . Obtaining a high information gain, therefore, implies achieving a significant reduction of entropy or uncertainty caused by considering the influence of that attribute.

2.3 Iterative rule base generation

A data-driven rule base learning mechanism intuitively extracts rules from raw data to generate a rule base, which are in the format of antecedents associated with a corresponding consequent (Wang and Mendel 1992; Hong and Lee 1996). Rule base generation can also follow an iterative procedure (Hoffmann 2004; Galea and Shen 2006) to incrementally add new rules to the rule base. This section outlines an iterative rule base generation procedure, which repeatedly sequentially extracts rules from data into an emerging rule base.

Given a set of instances which consist of r antecedent attributes and a consequent attribute, a rule base is generated in an iterative procedure as illustrated in Algorithm 2. Here, fuzzy rules are considered for generality, which may be readily degenerated into a crisp rule set if preferred. The iteration process is terminated by checking against a pre-set threshold value that determines at least how many data points have been covered by the extracted rules so far.

Before the iterative procedure is executed to generate the rule base, the domains of all r antecedent attributes and the consequent attribute are quantified evenly into m_1, m_2, \dots, m_r and m_c fuzzy regions, respectively, where m_c denotes the number of regions for the consequent

Algorithm 2 Iterative Rule Extraction

Input: Data set of instances D , threshold value δ

Output: Rule base R

```

1: Divide the domain of each antecedent and consequent attribute
   evenly into a certain number of fuzzy regions, and construct the
   fuzzy region space (FRS) of the antecedent, which is a hypercube
   with the dimensionality of  $m_1 \times m_2 \times \dots \times m_r$ , where  $m_i$ ,  $i = 1, \dots, r$ 
   stands for the number of regions for the  $i$ th attribute;
2: while true do
3:   Apply the data of instances  $D$  into the FRS and match each
   instance to a corresponding hypergrid in the hypercube in terms
   of its antecedent attributes;
4:   Select the hypergrid with the highest hits, denoting  $n$  as the highest
   hit number;
5:   if  $n > \delta$  then
6:     Extract a rule from this hypergrid, and add it into the rule base
      $R$ ;
7:     Remove all of the instances which hit this hypergrid from  $D$ ,
     update  $D = D - D_{\text{most\_hit}}$ ;
8:   else
9:     End While Loop;
10:  end if
11: end while
12: Return  $R$ 

```

attribute. Each fuzzy region is assigned with a membership function (implemented with triangular membership functions in this work for simplicity). This results in a division of fuzzy region space of the antecedent of an emerging rule in the form of a hypercube, of which each hypergrid stands for a combination of particular fuzzy regions of the r antecedent attributes.

The iteration process begins with the complete data set of instances D . A hypergrid hit by an instance indicates the largest value of membership is obtained for the corresponding combination of fuzzy regions. The hypergrid which is most covered by the instances in D receives the most hits amongst all. As indicated above, the threshold δ is used to determine whether the most covered hypergrid can form a rule and be added into the rule base R . If the number of the highest hits is larger than the threshold, a rule is extracted from this hypergrid.

The rule antecedent values returned by this iteration are those fuzzy values associated with the corresponding hypergrid. The rule consequent adopts the fuzzy value which corresponds to one of the m_c values at which the instances have the highest number of hits. After this, those instances hit in this hypergrid are removed from the original data set, and the iterative process repeats by treating the remaining data as the input data set to start the next round for the generation of the rules following the current one. However, if the proportion of hit instances is less than δ , a rule cannot be generated by this hypergrid because those small number of hits may just be due to noise, and the iterative procedure is hence terminated.

This simple iterative rule generation procedure will be used to learn a rule base to construct the inference system proposed in Sect. 3 (assuming no rules are provided by domain experts). If the generated rule base is dense, any standard fuzzy rule inference technique (e.g., compositional rule of inference (CRI)) can be employed to perform classification once a new input observation is provided. Otherwise, the observation is used as the input to the fuzzy rule interpolation process if it does not match any learned rules. Of course, if it matches a certain rule in the space rule base, CRI will be used as usual.

3 Antecedent weighted T-FRI

This section presents a novel technique for fuzzy rule interpolation which is guided with antecedent weights obtained by information gain. The proposed inference system is illustrated in Fig. 3. The iterative rule learning procedure presented in Sect. 2.3 generates the rule base from data. The scale and move transformation-based fuzzy rule interpolation (T-FRI) is utilised to work with information gain here. Note that the computation on information gain precedes, and its results are used for, all three key stages in T-FRI. The antecedent weighted T-FRI using information gains is described in the following with an illustrative example to show how it works.

3.1 Illustrative case

To illustrate the proposed work, a simple fuzzy classification problem (Yuan and Shaw 1995) is utilised here, involving a small set of training data of 16 instances. The system is set to make a decision on what sports activity to be undertaken (namely, volleyball, swimming and weight lifting) given the status of four conditional attributes regarding the weather, in terms of temperature (hot, mild and cool), outlook (sunny, cloudy and rain), humidity (humid and normal) and wind (windy and not windy).

Six fuzzy rules have been generated as given below. However, these six rules form a dense rule base where the domains of the antecedent variables are completely covered by the rules. To facilitate the illustration (of interpolation), Rule 6 is purposefully removed to have a sparse rule base.

1. If *Temperature* is *Hot* and *Outlook* is *Sunny*, then *Swimming*.
2. If *Temperature* is *Hot* and *Outlook* is *Cloudy*, then *Swimming*.
3. If *Outlook* is *Rain*, then *Weight lifting*.
4. If *Temperature* is *Mild* and *Wind* is *Windy*, then *Weight lifting*.

5. If *Temperature* is *Mild* and *Wind* is *Not Windy*, then *Volleyball*.
6. (If *Temperature* is *Cool*, then *Weight lifting*.)

3.2 Turning rules into training data via reverse engineering

Given a rule base, the proposed information gain-guided T-FRI begins with a reverse engineering procedure which converts the rules into a set of artificial training samples, forming a decision table for the calculation of required information gains. This development is based on the examination of how T-FRI performs its task. Its first key stage is the selection of n closest fuzzy rules when an observation is presented (which does not match with any existing rule in the sparse rule base and hence, CRI is not applicable).

In conventional T-FRI algorithms, all antecedent attributes of the rules are assumed to be of equal significance while searching for a subset of rules closest to the observation since the original approaches are unable to assess, nor to make use of, the relative importance or ranking of these antecedent attributes. Information gain offers such an intuitively sound and implementation-wise straightforward mechanism for evaluating the relative significance of attributes.

The question is what data are available to act as the learning examples for computing the information gains. T-FRI works with a sparse rule base. When an observation is given, it is expected to produce an interpolated result for the consequent variable. Without losing generality, it is practically presumed that there is no sufficient example data available for use to support the computation of the required information gains due to the sparseness of domain knowledge. However, any T-FRI method does use a given sparse rule base involving a set of antecedent variables $Y = A \cup \{z\}$ (as shown in Sect. 2.1). This set of rules can be translated into an artificial *decision table* (i.e., a set of artificially generated training examples), where each row represents a particular rule. In any data-driven learning mechanism, rules are learned from given data samples. Translating rules back to data is therefore a reverse engineering process of data-driven learning.

Generally speaking, a sparse rule-based system may involve rules that use different numbers of antecedent variables and even different variables in the first place. In order to employ the proposed reverse engineering procedure to obtain a training decision table, all rules are reformulated into a common representation by the following two-step procedure:

- Identifying all possible antecedent variables appearing in the rules and all value domains for these variables, and
- Expanding iteratively each existing rule into one which involves all domain variables such that if a certain antecedent variable is not originally involved in a rule,

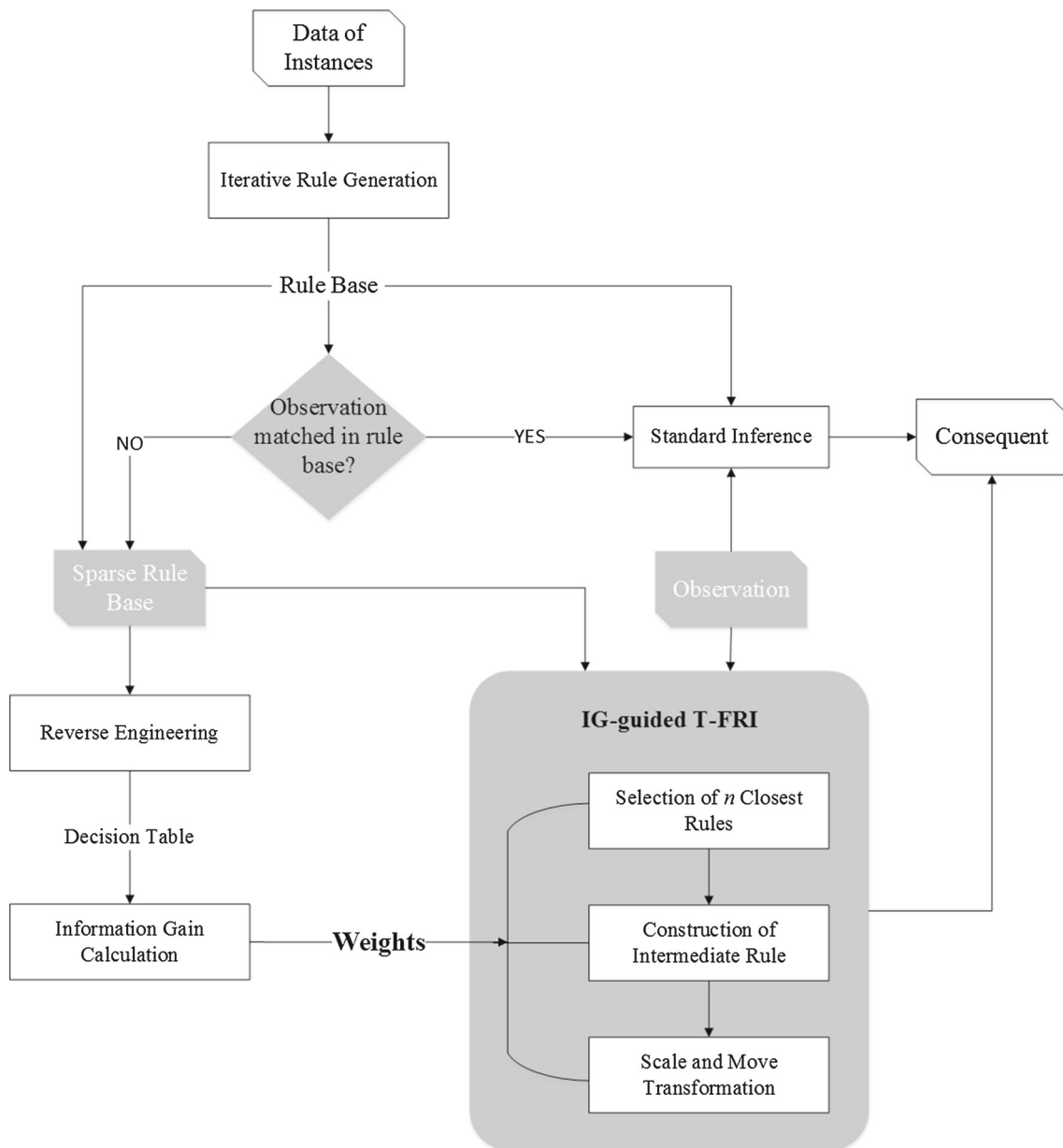


Fig. 3 Proposed inference system

then that rule is replaced by q rules, with q being the cardinality of the value domain of that variable, so that the variable within each of the expanded rule takes one possible and different value from its domain.

The above procedure makes logical sense. This is because for any rule, if a variable is missing from the rule antecedent, it means that it does not matter what value it takes and the rule will lead to the same consequent value, provided that those variables that do appear in the rule are satisfied.

Given the rule base of Sect. 3.1 which may be reformulated as given in Table 1. Following the two-step pro-

cedure, 32 training data are generated as listed in Table 9 in “Appendix A”. The reverse engineering process can be explained using the illustrative case. Without losing generality, assume that the first given rule is used to create the artificial data first. Then, part of the emerging artificial decision table is constructed from this rule first. Note that *Humidity* and *Wind* are missing in *Rule 1*, which means if *Temperature* is satisfied with the value *Hot* and *Outlook* with *Sunny*, the rule is satisfied and thus, the consequent variable *Decision* will have the value of *Swimming* no matter which value *Humidity* and *Wind* takes. That is, *Rule 1* can be expanded by the first four data in Table 9, each having

Table 1 Rule base in illustrative case

Rules	Variables				
	Temperature	Outlook	Humidity	Wind	Decision
r^1	<i>Hot</i>	<i>Sunny</i>	–	–	<i>Swimming</i>
r^2	<i>Hot</i>	<i>Cloudy</i>	–	–	<i>Swimming</i>
r^3	–	<i>Rain</i>	–	–	<i>Weight lifting</i>
r^4	<i>Mild</i>	–	–	<i>Windy</i>	<i>Weight lifting</i>
r^5	<i>Mild</i>	–	–	<i>Not windy</i>	<i>Volleyball</i>

the variable *Humidity* and *Wind* taking one of its two possible values. Similarly, more artificial data can be created by translating and expanding the remaining original rules.

Comparing both the antecedent values and the consequent in Table 9, it can be seen that there are several identical samples which are generated from different original rules. Retaining one of them results in a total of 30 training data. Note that in such an artificially constructed decision table, it may appear to include inconsistent data since they may have the same values for the respective antecedent attributes but different consequents (e.g., two inconsistent pairs are italicised in Table 9). This does not matter as the eventual rule-based inference, including rule interpolation does not use these artificially generated rules, but the original sparse rule base. They are created just to help assess the relevant significance of individual variables through the estimation of their respective information gains. It is because there are variables which may lead to potentially inconsistent implications in a given problem that it is possible to distinguish the different abilities of the variables in possessing the power in influencing the consequent. This in turn enables the measuring of the information gains of individual antecedent variables as described below.

3.3 Weighting of individual variables

Given an artificial decision table that is derived from a sparse rule base via reverse engineering, the information gain IG'_i of a certain antecedent variable a_i , $i = 1, \dots, m$, regarding the consequent variable z is calculated as per Eq. (6):

$$IG'_i = \text{Entropy}(\{z\}) - \sum_{v \in \text{Value}(a_i)} \frac{|\{z\}_v|}{|\{z\}|} \text{Entropy}(\{z\}_v) \quad (7)$$

where $\{z\}_v$ denotes the subset of rules in the artificial decision table in which the antecedent variable a_i has the value v . Repeating the above, the information gains for all antecedent variables IG'_i , $i = 1, \dots, m$ can be computed. These values are then normalised into IG_i , $i = 1, \dots, m$ such that

$$IG_i = \frac{IG'_i}{\sum_{t=1, \dots, m} IG'_t} \quad (8)$$

Table 2 Weighted decision table with information gain calculated for each antecedent variable

Rules	Variables				
	a_1	a_2	\dots	a_m	z
r^1	A_1^1	A_2^1	\dots	A_m^1	z^1
r^2	A_1^2	A_2^2	\dots	A_m^2	z^2
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
r^N	A_1^N	A_2^N	\dots	A_m^N	z^N
Weight	IG_1	IG_2	\dots	IG_m	

Table 3 Normalised information gains calculated using 30 training samples

Antecedent	<i>Temperature</i>	<i>Outlook</i>	<i>Humidity</i>	<i>Wind</i>
Normalised IG	0.5000	0.4515	0.0000	0.0485

Given the inherent meaning of information gain, the resulting normalised values can be intuitively interpreted as the relative significance degrees of the individual rule antecedent attributes in the determination of the rule consequent. Therefore, they can be used to act as the weights associated with each individual antecedent variable in the original sparse rule base. In general, through this procedure, an original decision table such as the one shown in Table 1 becomes Table 2 (where N is the number of the distinct rules generated by the procedure), with a weight added to each antecedent variable.

Recall the example case. The normalised information gains calculated for each antecedent variable using those 30 training samples are shown in Table 3. The information gain of the antecedent attribute *Temperature* is relatively higher than the other three, which indicates *Temperature* plays a much important role in the decision on the sports activity. This can be verified from the five fuzzy rules where the antecedent variable *Temperature* appears in 4 rules. On the other hand, *Humidity* and *Wind* are assigned a very small amount of weight. In particular, the normalised IG of *Humidity* is 0, signifying its irrelevance on the decision in this rule base.

3.4 Weighted T-FRI

Given the weights associated with the rule antecedent attributes T-FRI can be modified. Such modification will involve three key stages as detailed below.

3.4.1 Weight-guided selection of n closest rules

First of all, when an observation is present which does not entail a direct match with any rule in the sparse rule base, n ($n \geq 2$) closest rules to it are required to be chosen to perform rule interpolation. The original selection is based on the Euclidean distance measured by aggregating the distances between individual antecedent variables of a certain rule and the corresponding variable values in the observation [as per Eq. (3)]. Considering the weights assessed by information gain, the distance between a given rule r^p and the observation o^* can now be calculated by

$$\tilde{d}(r^p, o^*) = \frac{1}{\sum_{t=1}^m IG_t^2} \sqrt{\sum_{j=1}^m (IG_j d(A_j^p, A_j^*))^2} \quad (9)$$

where $d(A_j^p, A_j^*)$ is computed according to Eq. (4).

Choosing the n closest rules this way allows those rules which involve certain antecedent variables that are regarded more significant to be selected with priority. Note that the normalisation term $\frac{1}{\sum_{t=1}^m IG_t^2}$ is a constant and, therefore, can be omitted in computation since the purpose of calculating the distance $\tilde{d}(r^p, o^*)$ is in order to rank the rules and only information on the relative distance measures is required.

To continue illustration with the case study, suppose that the membership functions used to describe the antecedent and consequent variables are defined as given in Fig. 5 of “Appendix B”. Also, suppose that the observation of Table 4 (involving only singleton fuzzy sets) is presented, resulting in the membership values for the observation as shown in the bottom of row of Table 4. This does not match with any of the rules in the sparse rule base. Thus, no rule in the sparse rule base can be fired directly and FRI is applied to derive a conclusion. Both the information gain-guided T-FRI (IG-T-FRI) and the original T-FRI are employed here for comparison. Given the rule base and the observation, the 2 closest rules selected by T-FRI and those by IG-T-FRI are different, with

Rules 4 and 5 and Rules 3 and 5 are selected by T-FRI and IG-T-FRI, respectively.

3.4.2 Weighted parameters for intermediate-rule construction

Unlike the conventional T-FRI, the significance of individual antecedent variables is captured and reflected in their contribution towards the derivation of the (interpolated) consequent, by the use of their associated weights. To emphasise this, weights are integrated in all calculations during the transformation process, including the initial construction of the intermediate rule. In particular, the weighting on the consequent \tilde{w}_z^i is now computed as follows:

$$\tilde{w}_z^i = \sum_{j=1}^m IG_j w_j^i \quad (10)$$

This is a direct extension to the original construction process of the intermediate rule as shown in step 5 of Algorithm 1 where all variables are equally regarded in terms of their significance. Referring to Eq. (8), it is clear that if antecedent attributes are of equal significance, \tilde{w}_z^i degenerates to w_z^i .

3.4.3 Weighted transformation

In performing the scale and move transformations, the previous computation of the required scale and move factors, namely those equations in step 11 of Algorithm 1, is now modified to:

$$\tilde{s}_z = \sum_{j=1}^m IG_j s_{A_j}, \quad \tilde{m}_z = \sum_{j=1}^m IG_j m_{A_j} \quad (11)$$

From these modifications, given an observation (that does not match with any rule in the sparse rule base), an interpolated consequent variable \tilde{z}^* can be obtained by performing the transformation $T(\tilde{z}', \tilde{s}_z, \tilde{m}_z)$. Note that when all weights are equal, i.e., when all antecedent variables are assumed to be of equal significance, the above modified version degenerates to the original T-FRI. Mathematical proof for this is straightforward and, hence, omitted here.

Table 4 Observation in illustrative example

Antecedent attribute	Temperature			Outlook			Humidity		Wind	
Observed value	0.91			0.42			0.5		0.51	
Membership value	<i>Hot</i>	<i>Mild</i>	<i>Cool</i>	<i>Sunny</i>	<i>Cloudy</i>	<i>Rain</i>	<i>Humid</i>	<i>Normal</i>	<i>Windy</i>	<i>Not windy</i>
	0.0	0.0	0.775	0.0	0.733	0.0	0.5	0.5	0.49	0.51

Return to the illustrative case, applying the above improved T-FRI with weighted parameters to the example leads to the following intermediate rule using *Rules 3* and *5*:

If *Temperature* is (0.78,0.91,1.03) and *Outlook* is (0.31,0.47,0.47) and *Humidity* is (0.50,0.50,0.50) and *Wind* is (0.20,0.66,0.66), then *Decision* is (2.49,2.49,2.49)

Differently, the intermediate rule created by the two closest rules, *Rules 4* and *5* using T-FRI is:

If *Temperature* is (0.61,0.91,1.21) and *Outlook* is (0.42,0.42,0.42) and *Humidity* is (0.50,0.50,0.50) and *Wind* is (0.01,0.51,1.01), then *Decision* is (2.51,2.51,2.51)

Given the simplified case where observations are all singleton fuzzy sets, the above intermediate results imply that the final interpolated result with IG-T-FRI is $\tilde{z}^* = (2.49, 2.49, 2.49)$, using the IG-guided transformation $T(\tilde{z}') = (2.49, 2.49, 2.49)$, $\tilde{s}_z = 0$, $\tilde{m}_z = 0$, and that the result with the standard T-FRI is $z^* = (2.51, 2.51, 2.51)$, using a transformation of $T(z') = (2.51, 2.51, 2.51)$, $s_z = 0$, $m_z = 0$. From this, through defuzzification (to obtain a classification result), the conclusions drawn by the use of these two different methods are *Weight lifting* and playing *Volleyball*, respectively. Clearly, the outcome of applying IG-T-FRI has a better intuitive appeal given the particular observation. Indeed, recall the original rule base for this illustrative case given in (Yuan and Shaw 1995), the observation used for illustration actually matches *Rule 6* (i.e., the one purposefully removed to form a sparse rule base). This results in the same decision if fired as the interpolated consequent derived by the proposed IG-T-FRI method.

The workflow of the construction of the intermediate rule and of the computation of the interpolative results for both methods is outlined in Fig. 6 in “Appendix C”.

This illustrative case is very simple, involving only a small number of instances and a rather specific rule base. It is therefore not surprising that similar interpolated values may result by the use of either the original T-FRI or the proposed IG-T-FRI. Even though the above still demonstrates the strength of the proposed approach, the following section will systematically evaluate such strength using more complicated datasets.

4 Experimental evaluation

This section presents a systematic experimental evaluation of the proposed inference system, where the information gain-guided T-FRI approach is embedded. The work is assessed for the task of performing pattern classification over nine benchmark datasets. Classification results are compared with those obtained by the original T-FRI method and also, with the standard Mamdani inference (Mamdani and Assilian 1999) without involving rule interpolation but directly firing

Table 5 Datasets used

Dataset	Attributes #	Classes #	Instances #
Iris	4	3	150
Diabetes	8	2	768
Phoneme	5	2	5404
Appendicitis	7	2	106
Magic	10	2	1902
NewThyroid	5	3	215
Banana	2	2	5300
Haberman	3	2	306
Monk-2	6	2	432

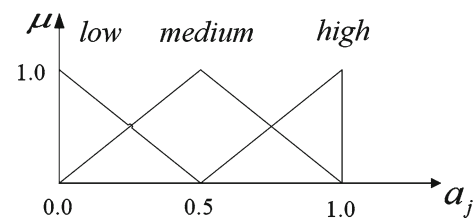


Fig. 4 Membership functions defining the linguistic terms

those (possibly partially) matched rules. In addition, a statistical analysis is utilised to further evaluate the performance of the proposed approach over the original T-FRI.

4.1 Experimental set-up

4.1.1 Datasets

The nine benchmark datasets are taken from the UCI machine learning (A Asuncion 2007) and KEEL (Knowledge Extraction based on Evolutionary Learning) (Alcalá et al. 2010) dataset repositories, with their details summarised in Table 5.

4.1.2 Experimental methodology

Triangular membership functions are used to represent the fuzzy sets of the antecedent variables due to their popularity and simplicity. Given that the problems are all for classification, the consequent variable always adopts a singleton fuzzy set (i.e., a crisp value) as its value. In general, different variables have their own underlying domains. However, to simplify knowledge representation, these domains are normalised to take a value from the common range of 0 to 1, as illustrated in Fig. 4. Note that such a simple fuzzification is used in this work; no optimisation of the value domain is carried out. This is used for all methods under comparison. A fine-tuned definition of the membership functions will no doubt improve the performance of the classification results.

Experiments are validated by tenfold cross-validation which is repeated for 10 times per dataset. The rule base

is generated from training data after the fuzzification by the presented iterative rule induction method. In particular, the domain interval of antecedent variables is divided into three fuzzy regions, and the threshold δ is empirically set to 2 in order to determine whether to promote a rule into the emerging rule base. 10% of the learned rules are deliberately removed randomly, ensuring that the resultant rule base is sparse. The information gains for weighting are computed using an artificial decision table which is translated from the learned rule base. The number of the closest rules to perform rule interpolation is set to 2, which is commonly used in the existing literature. The classification performance is assessed in terms of accuracy over the testing data. A statistical t test ($p = 0.05$) is utilised to determine the statistical significance of the improvement of the information gain-guided T-FRI over the original T-FRI for each of the nine data set.

4.2 Results and discussion

4.2.1 Comparison on overall classification accuracy

Table 6 shows the classification performance over the nine datasets, measured with the average accuracy and the standard deviation (SD) through a process of 10×10 cross-validation. In particular, the column of *CRI* presents the results obtained using the compositional rule of inference directly by firing those matched rules only; the *T-FRI* column shows the results obtained by the use of the original T-FRI; and the *IG-T-FRI* column summaries the results obtained using the information gain-guided T-FRI approach. A pairwise t test ($p = 0.05$) validates the experimental evaluation furthermore. Note that the asterisk (*) after a result in the column *T-FRI* indicates that the improvement made by the original T-FRI over *CRI* is statistically significant, and similarly the asterisk (*) in the *IG-T-FRI* column shows that the improvement made by *IG-T-FRI* is in turn, statistically significant over T-FRI.

Table 6 Average classification accuracy (%) and standard deviation with 10×10 fold cross-validation

Dataset	<i>CRI</i>	<i>T-FRI</i>	<i>IG-T-FRI</i>
Iris	66.66 \pm 0.25	76.99 \pm 0.16*	82.53 \pm 0.13*
Diabetes	32.10 \pm 0.08	62.50 \pm 0.06*	68.49 \pm 0.05*
Phoneme	38.40 \pm 0.09	60.53 \pm 0.05*	66.18 \pm 0.07*
Appendicitis	32.27 \pm 0.10	57.72 \pm 0.12*	69.69 \pm 0.13*
Magic	49.15 \pm 0.05	58.40 \pm 0.09*	64.67 \pm 0.05*
NewThyroid	43.33 \pm 0.28	47.43 \pm 0.24*	53.28 \pm 0.22*
Banana	44.83 \pm 0.08	60.49 \pm 0.05*	63.27 \pm 0.04*
Haberman	54.00 \pm 0.09	71.73 \pm 0.08*	77.47 \pm 0.07*
Monk-2	32.63 \pm 0.05	60.01 \pm 0.11*	63.31 \pm 0.06*
Average	43.70 \pm 0.12	61.75 \pm 0.11	67.65 \pm 0.09

The accuracies achieved by *CRI* indicate the sparseness of the rule base, which is expected to have a relatively poorer performance in classification. Both the interpolative reasoning approaches clearly show their significant advantage in dealing with the sparse rule base. Importantly, the information gain-guided T-FRI method has consistently achieved better classification accuracies over all nine datasets, with an overall 5.9% higher accuracy than that reachable by the original T-FRI and, a 23.95% improvement over the use of *CRI* which only fires matched rules without any rule interpolation method. The SD values compared between the three methods indicate that more robust classification performance is achieved by *IG-T-FRI* also. Together, these results clearly demonstrate the potential of the proposed work.

4.2.2 Comparison on false negatives and false positives

Apart from the classification accuracies, in many real-world applications, it is worth to examine the statistical rates on *true positives* (TP), *true negatives* (TN), *false positives* (FP) and *false negatives* (FN). Without overly complicating the experimental investigation while having a focused discussion, the *Diabetes* dataset, as a binary classification problem, is selected for this comparison. Tables 7 and 8 show the confusion matrices computed by the use of the original T-FRI and that of *IG-T-FRI*, respectively. ‘Positive’ in both tables is interpreted as an instance in which a person is diagnosed to have diabetes. The numbers shown in both tables are computed by averaging the results obtained in 10×10 cross-validation.

First of all, recall the results shown in Table 6, the classification accuracy of T-FRI is 62.5%, which is improved to 68.49% using *IG-T-FRI*. As can be seen from comparing Tables 7 and 8, as the classification precision increases with

Table 7 Confusion matrix of T-FRI on *diabetes* dataset by averaging 10×10 cross-validation

	Classified	
	Positive	Negative
Actual		
Positive	9.5	17.6
Negative	11.2	38.5

Table 8 Confusion matrix of *IG-T-FRI* on *Diabetes* dataset by averaging 10×10 cross-validation

	Classified	
	Positive	Negative
Actual		
Positive	17.2	9.9
Negative	14.3	35.4

the use of IG-T-FRI, the rate of FN reduces significantly from 64.94 to 36.53% [where the false negative rate is calculated by $FN/(TP + FN)$]. This makes a great sense in performing medical diagnosis since the rate of missing disease detection (i.e., the proportion of the disease tested as not present when it is really present) is reduced. Although the number of FP is slightly increased, the diagnostic sensitivity (true positive rate) has raised significantly also, with 28.41% in average. This promising result clearly indicates considerable improvement on the decisions made by the use of IG-T-FRI.

5 Conclusion

This paper has presented a novel fuzzy rule-based inference system to address the situation when the rule base is sparse. The proposed information gain-guided fuzzy rule interpolation approach is embedded in this system, where the rule antecedent variables are weighted via computing the information gains. In particular, the computation is enabled through an innovative reverse engineering procedure which converts fuzzy rules into training samples. The proposed method is illustrated by a case study with a small data set and is systematically evaluated by solving benchmark classification problems over nine datasets. The experimental results have confirmed that the relative significance of the individual rule antecedent variables can indeed be captured by the information gains, forming the weights on the variables to guide FRI. This remarkably improves the performance of the interpolative reasoning, thanks to the exploitation of the information gains in differentiating the significances of different antecedent variables.

While very promising, much can be done to further improve this proposed work. The present implementation assumes the use of a data-driven rule learning mechanism that converts a given dataset into rules, with a simple fuzzification procedure. The size of the rule base may be very large due to a large dataset. Any other rule induction techniques (e.g., those reported in (Janikow 1998; Afify 2016)) that may be used as an alternative to generate a more compact rule base would be helpful, improving the performance of the interpolation method further. With the introduction of information gain

in support of weighted rule interpolation, there may be an additional computation overhead overall as compared to the use of the original T-FRI algorithm. An experimental analysis of the runtime expense, in comparison with T-FRI, forms another piece of interesting further work. Finally, the current approach assumes a fixed (sparse) rule base. However, having run the process of rule interpolation, intermediate fuzzy rules are generated. These can be collected and refined to form additional rules to support subsequent inference, thereby enriching the rule base and avoiding unnecessary interpolation afterwards (Naik et al. 2017).

Acknowledgements The first author is grateful to the China Scholarship Council and Aberystwyth University for their support in this research. The authors would like to thank the reviewers of the original version of this paper that was presented at the 16th UK Workshop on Computational Intelligence, 2016; their constructive comments have helped improve this work significantly, leading to it receiving one of the two best paper awards at the Workshop.

Funding This study was partly funded by the National Key Research and Development Program of China (Grant No. 2016YFB0502502).

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix A: Reverse engineering in illustrative example

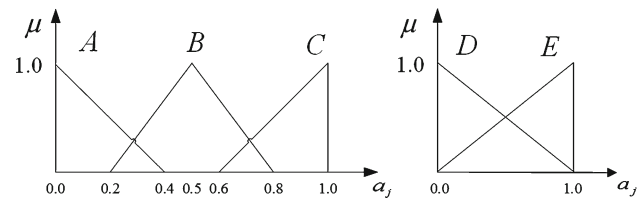
The artificial decision table converted from the five fuzzy rules by reverse engineering in the illustrative case is shown in Table 9, where two pairs of *inconsistent* data are italicised.

Table 9 32 Training data after reverse engineering given five rules

Temperature	Outlook	Humidity	Wind	Decision
Hot	Sunny	Humid	Windy	Swimming
Hot	Sunny	Humid	Not Windy	Swimming
Hot	Sunny	Normal	Windy	Swimming
Hot	Sunny	Normal	Not Windy	Swimming
Hot	Cloudy	Humid	Windy	Swimming
Hot	Cloudy	Humid	Not Windy	Swimming
Hot	Cloudy	Normal	Windy	Swimming
Hot	Cloudy	Normal	Not Windy	Swimming
Hot	Rain	Humid	Windy	Weight lifting
Hot	Rain	Humid	Not Windy	Weight lifting
Hot	Rain	Normal	Windy	Weight lifting
Hot	Rain	Normal	Not Windy	Weight lifting
Mild	Rain	Humid	Windy	Weight lifting
Mild	Rain	Humid	Not Windy	Weight lifting
Mild	Rain	Normal	Windy	Weight lifting
Mild	Rain	Normal	Not Windy	Weight lifting
Cool	Rain	Humid	Windy	Weight lifting
Cool	Rain	Humid	Not Windy	Weight lifting
Cool	Rain	Normal	Windy	Weight lifting
Cool	Rain	Normal	Not Windy	Weight lifting
Mild	Sunny	Humid	Windy	Weight lifting
Mild	Sunny	Normal	Windy	Weight lifting
Mild	Cloudy	Humid	Windy	Weight lifting
Mild	Cloudy	Normal	Windy	Weight lifting
Mild	Rain	Humid	Windy	Weight lifting
Mild	Rain	Normal	Windy	Weight lifting
Mild	Sunny	Humid	Not Windy	Volleyball
Mild	Sunny	Normal	Not Windy	Volleyball
Mild	Cloudy	Humid	Not Windy	Volleyball
Mild	Cloudy	Normal	Not Windy	Volleyball
Mild	Rain	Humid	Not Windy	Volleyball
Mild	Rain	Normal	Not Windy	Volleyball

Appendix B: Membership functions used in illustrative example

The triangular membership functions adopted in the illustrative case are shown in Fig. 5, which are used to represent all the antecedent variables for the original data set given in (Yuan and Shaw 1995). In particular, the variables *Outlook* and *Temperature* adopt the membership functions defined on the left of the figure, where *A*, *B*, *C* stand for *Sunny*, *Cloudy*, *Rain*, or *Hot*, *Mild*, *Cool*, respectively, in relation to the two variables. *Humidity* and *Wind* adopt the membership functions on the right, with *D*, *E* standing for *Humid*, *Normal* or *Windy*, *Not Windy*, respectively. Since this is a classification

**Fig. 5** Definition of linguistic terms for domain variables

problem, the consequent generally adopts the singleton fuzzy set in the description of the decision.

Appendix C: Workflow of illustrative example

The workflow of the construction of the intermediate rule and of the computation of the interpolative results using the original T-FRI or the information gain-guided T-FRI is shown in Fig. 6. The left hand side illustrates the workflow of the T-FRI and the right does that of IG-T-FRI. In particular, the fuzzy sets of the antecedent variables taken by the selected closest rules, observation and the intermediate rule are shown in the first row, while the consequent variable of those are displayed in the last. The observation of each antecedent variable and the consequent of the selected rules are both illustrated using singleton fuzzy sets for simplicity.

For T-FRI on the left of this figure, the fuzzy sets in dashed lines represent the variable values in *Rule 4*, while those in dash-dotted lines represent the sets in *Rule 5*. For IG-T-FRI on the right, the fuzzy sets in dashed lines represent the variable values in *Rule 3*, and those in dash-dotted lines represent the sets in *Rule 5*. The dotted lines represent the (fuzzy) values of the computed intermediate antecedent and consequent variables.

Note that in general and also, as in this simple case, a certain antecedent variable may not be present in both selected rules (e.g., *Temperature* and *Wind* are involved in *Rule 5* but missing in *Rule 3*). In such situations, the corresponding value of the given observation is employed to replace the missing one in the closest rules, facilitating the interpolation. This makes logical sense as the missing value of an antecedent variable in a rule indicates that any value in its domain may be matched, so the observation naturally provides the best replacement guided by the representative value. Note also that since the transformation factors in this very simple illustrative case are zero, the interpolated consequents are also shown in dotted lines in the last row of the figure.

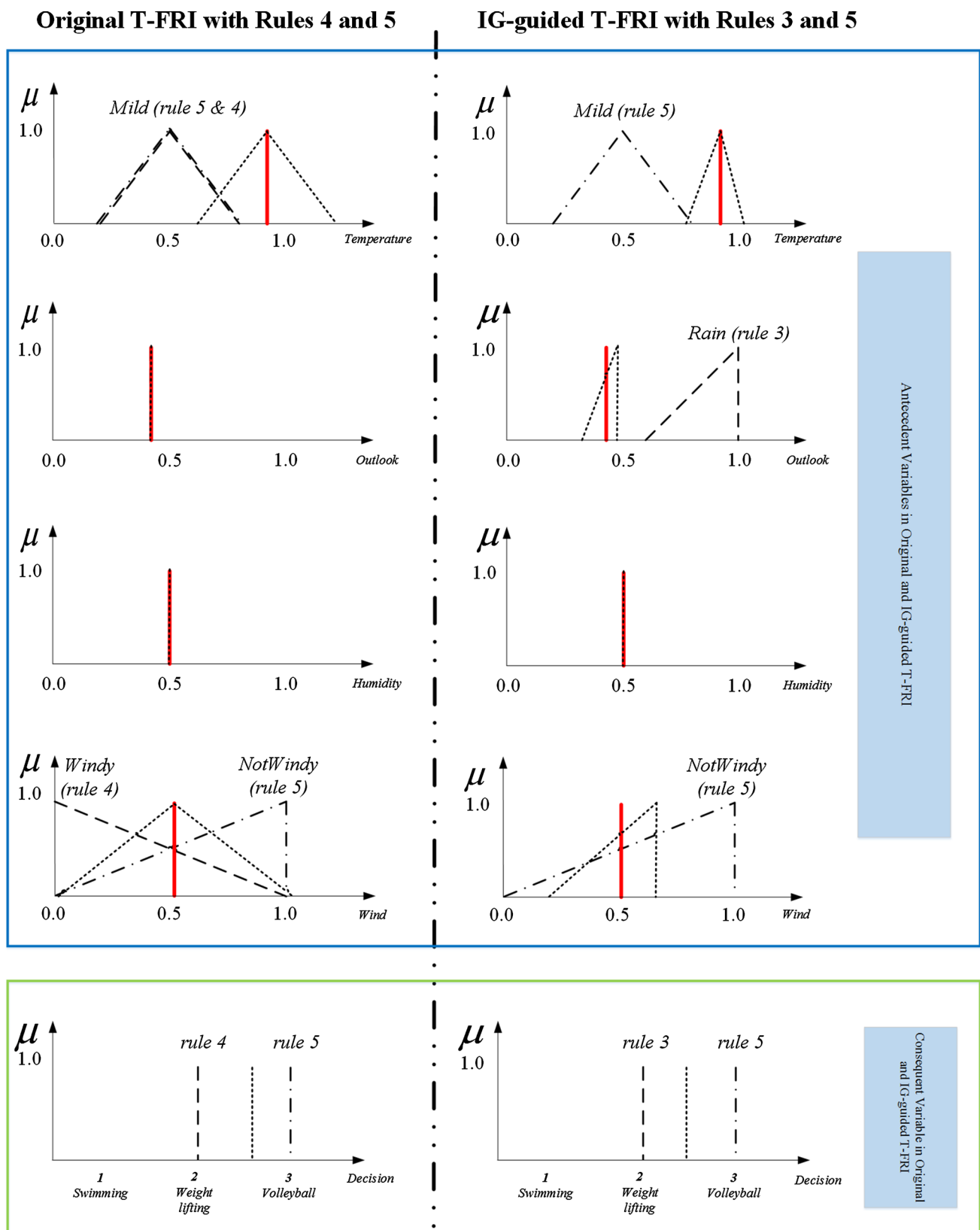


Fig. 6 Workflow of illustrative example

References

- A Asuncion DN (2007) UCI machine learning repository. <https://archive.ics.uci.edu/ml/datasets.html>
- Afify AA (2016) A fuzzy rule induction algorithm for discovering classification rules. *J Intell Fuzzy Syst* 30(6):3067–3085
- Alcalá J, Fernández A, Luengo J, Derrac J, García S, Sánchez L, Herrera F (2010) Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *J Mult Valued Logic Soft Comput* 17(2–3):255–287
- Chang YC, Chen SM, Liao CJ (2008) Fuzzy interpolative reasoning for sparse fuzzy-rule-based systems based on the areas of fuzzy sets. *IEEE Trans Fuzzy Syst* 16(5):1285–1301
- Chen SM, Chang YC (2011) Weighted fuzzy rule interpolation based on GA-based weight-learning techniques. *IEEE Trans Fuzzy Syst* 19(4):729–744
- Diao R, Jin S, Shen Q (2014) Antecedent selection in fuzzy rule interpolation using feature selection techniques. In: 2014 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE, pp 2206–2213
- Galea M, Shen Q (2006) Simultaneous ant colony optimization algorithms for learning linguistic fuzzy rules. In: Aghraham A, Grosan C, Ramos V (eds) *Swarm intelligence in data mining*. Springer, Berlin, pp 75–99
- Hoffmann F (2004) Combining boosting and evolutionary algorithms for learning of fuzzy classification rules. *Fuzzy Sets Syst* 141(1):47–58
- Hong TP, Lee CY (1996) Induction of fuzzy rules and membership functions from training examples. *Fuzzy Sets Syst* 84(1):33–47
- Hsiao WH, Chen SM, Lee CH (1998) A new interpolative reasoning method in sparse rule-based systems. *Fuzzy Sets Syst* 93(1):17–22
- Huang Z, Shen Q (2006) Fuzzy interpolative reasoning via scale and move transformations. *IEEE Trans Fuzzy Syst* 14(2):340–359
- Huang Z, Shen Q (2008) Fuzzy interpolation and extrapolation: a practical approach. *IEEE Trans Fuzzy Syst* 16(1):13–28
- Janikow CZ (1998) Fuzzy decision trees: issues and methods. *IEEE Trans Syst Man Cybern Part B (Cybern)* 28(1):1–14
- Jin S, Diao R, Quek C, Shen Q (2014) Backward fuzzy rule interpolation. *IEEE Trans Fuzzy Syst* 22(6):1682–1698
- Kóczy L, Hirota K (1993a) Approximate reasoning by linear rule interpolation and general approximation. *Int J Approx Reason* 9(3):197–225
- Kóczy L, Hirota K (1993b) Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases. *Inf Sci* 71(1–2):169–201
- Li YM, Huang DM, Zhang LN, et al (2005) Weighted fuzzy interpolative reasoning method. In: *Proceedings of 2005 international conference on machine learning and cybernetics, 2005*, vol 5. IEEE, pp 3104–3108
- Mamdani E, Assilian S (1999) An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Hum Comput Stud* 51(2):135–147
- Mitchell TM (1997) *Machine learning*. McGraw-Hill Science/Engineering/Math
- Naik N, Diao R, Shen Q (2017) Dynamic fuzzy rule interpolation and its application to intrusion detection. *IEEE Trans Fuzzy Syst*
- Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1(1):81–106
- Shannon CE (2001) A mathematical theory of communication. *ACM SIGMOBILE Mob Comput Commun Rev* 5(1):3–55
- Wang LX, Mendel JM (1992) Generating fuzzy rules by learning from examples. *IEEE Trans Syst Man Cybern* 22(6):1414–1427
- Yang L, Shen Q (2011) Adaptive fuzzy interpolation. *IEEE Trans Fuzzy Syst* 19(6):1107–1126
- Yang L, Chao F, Shen Q (2017) Generalized adaptive fuzzy rule interpolation. *IEEE Trans Fuzzy Syst* 25(4):839–853
- Yuan Y, Shaw MJ (1995) Induction of fuzzy decision trees. *Fuzzy Sets Syst* 69(2):125–139
- Zadeh L (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans Syst Man Cybern* 3:28–44
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353